

# Klassifizierung von Gefahrstoffen durch laserinduzierte Fluoreszenz (LIF) in der Ferndetektion

DWT-Tagung, Bonn 22. Februar 2018

Dr. Arne Walter, DLR Lampoldshausen



Wissen für Morgen



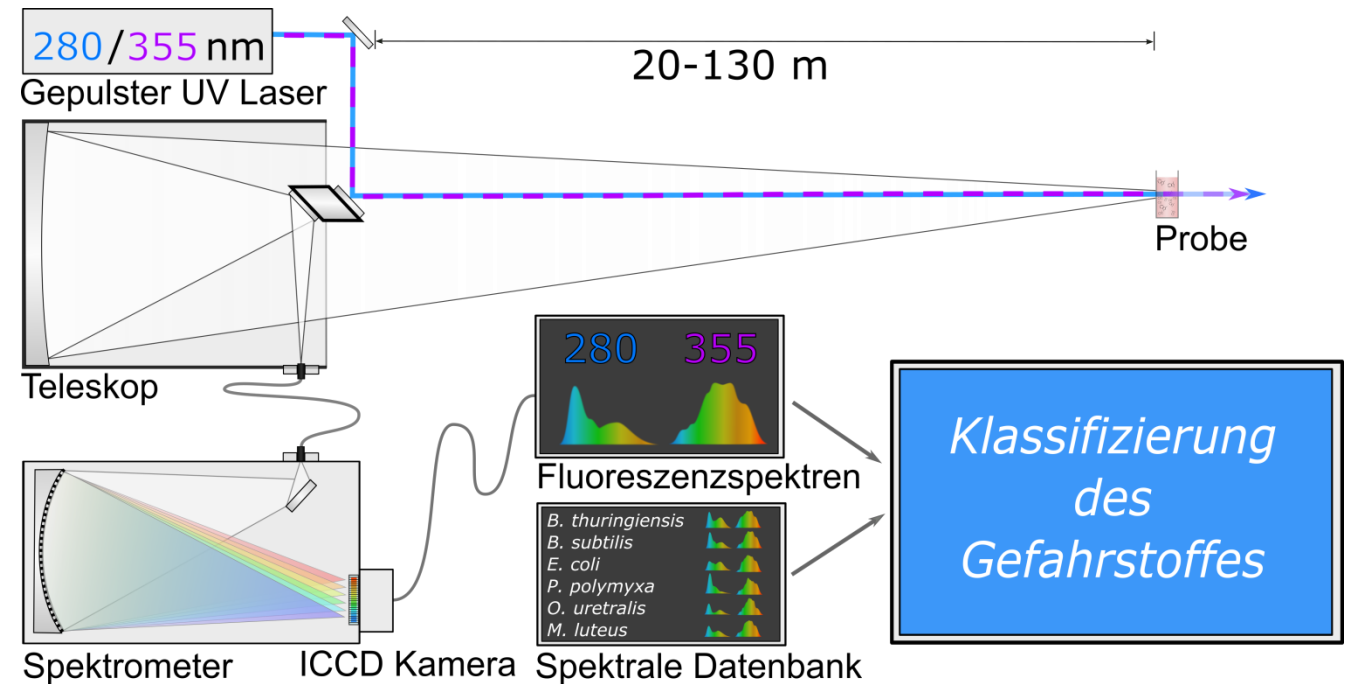
# Aufnahme der Spektren

## Freistrahlstrecke Lampoldshausen

- Realistische Messbedingungen
- 20 – 130 m (hier 22 m)

## STABIL Lasersystem

- 40 cm Teleskop
- Spektrometer
  - Spektrale Auflösung 1 nm
  - Bandbreite 300-600 nm
- Laser
  - Anregungswellenlänge 280 und 355 nm
  - Repetitionsrate 10 Hz, alternierend
  - Pulsenergie ~ 10 mJ
  - Pulsweite 7 ns



# Untersuchte Substanzen

Herbizide Pestizide	Lösungsmittel	Mineralöl	Pflanzenöl	Pflanzenstoff	Bakterien
Isoproturon Terbuthylazin Permethrin	Benzaldehyd Diethylether Isopropanol Terpentinersatz	Diesel (Aral) Diesel (HEM) Kerosin Motoröl (2 Takt) Paraffin	Rapsöl Sonnenblumenöl Kokosöl	Chlorophyll A beta-Carotin Lycopin Lutein Pollen <ul style="list-style-type: none"> <li>• Schwarzpappel</li> <li>• Sommer ESP</li> <li>• Sommer GER</li> <li>• Zitterpappel</li> </ul>	B. atrophaeus B. brevis B. subtilis B. thuringiensis E. coli O. urethralis P. polymyxa M. luteus



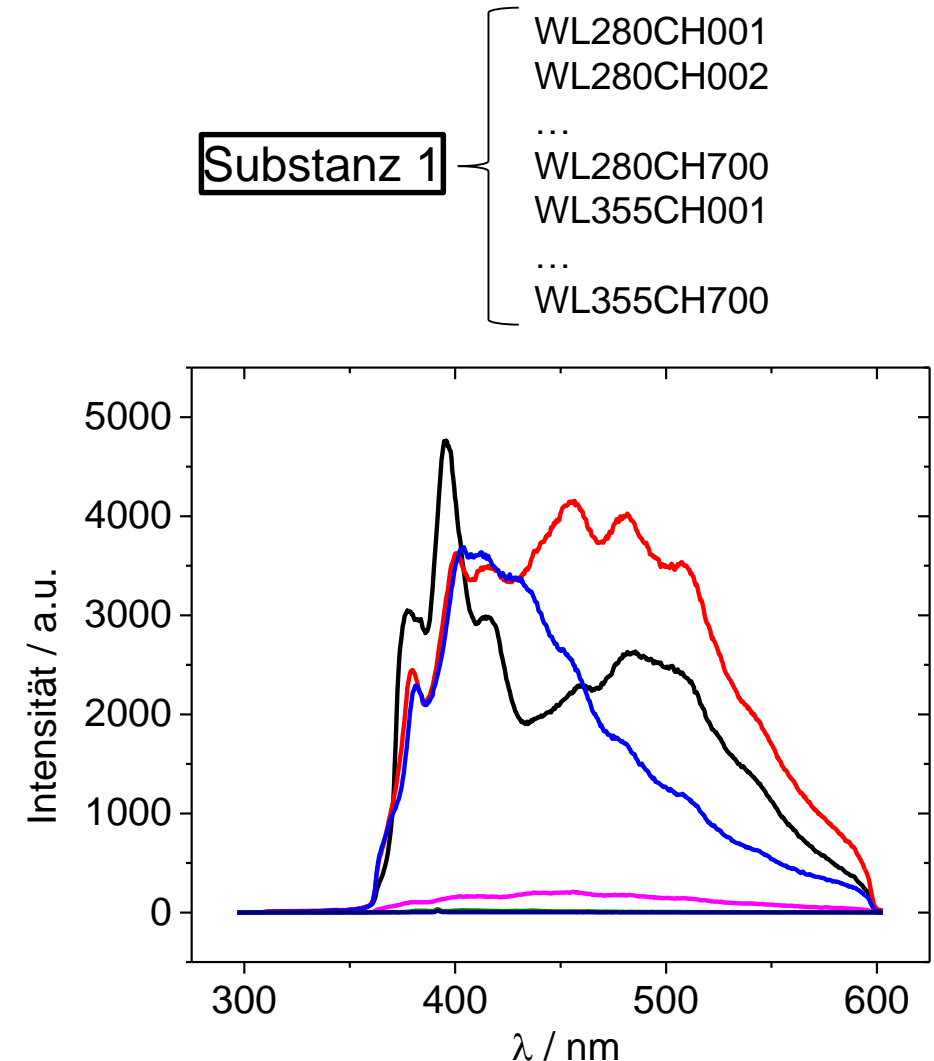
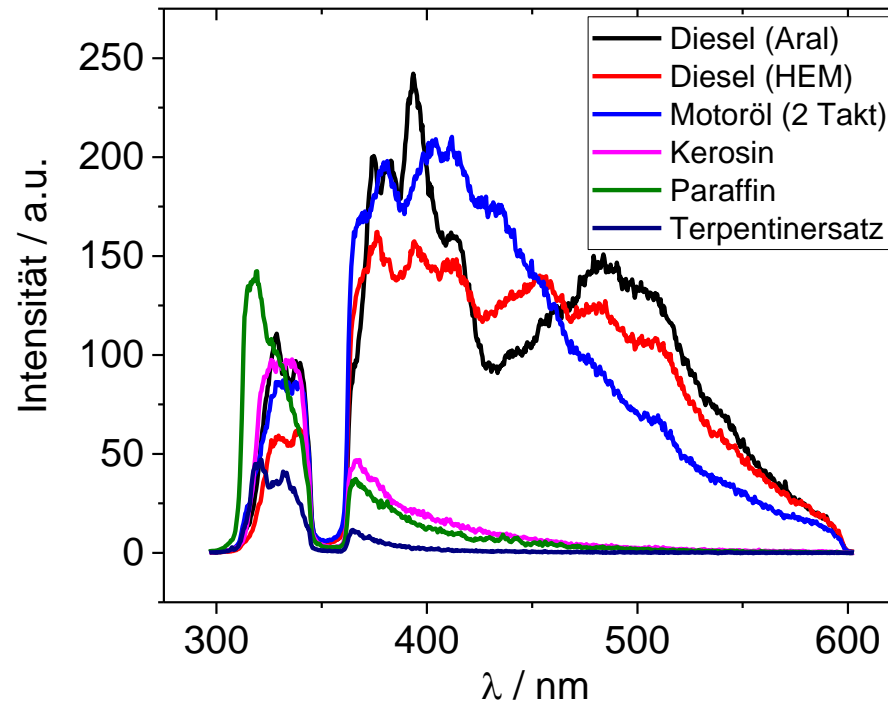
# Struktur der aufgenommenen Daten

## Datensatz pro Substanz

- 500 Einzelmessungen pro Anregungswellenlänge
- 720 Kanäle pro Spektrum

### Mineralöl

Diesel (Aral)  
Diesel (HEM)  
Motoröl (2 Takt)  
Kerosin  
Paraffin  
Terpentinersatz





# Einteilung in Substanzklassen

## 1. Zuordnung

- Einteilung der Substanzen in beliebige Klassen
  - Funktionell gefährlich / ungefährlich
  - Typ Bakterien / Chemikalien
  - ...

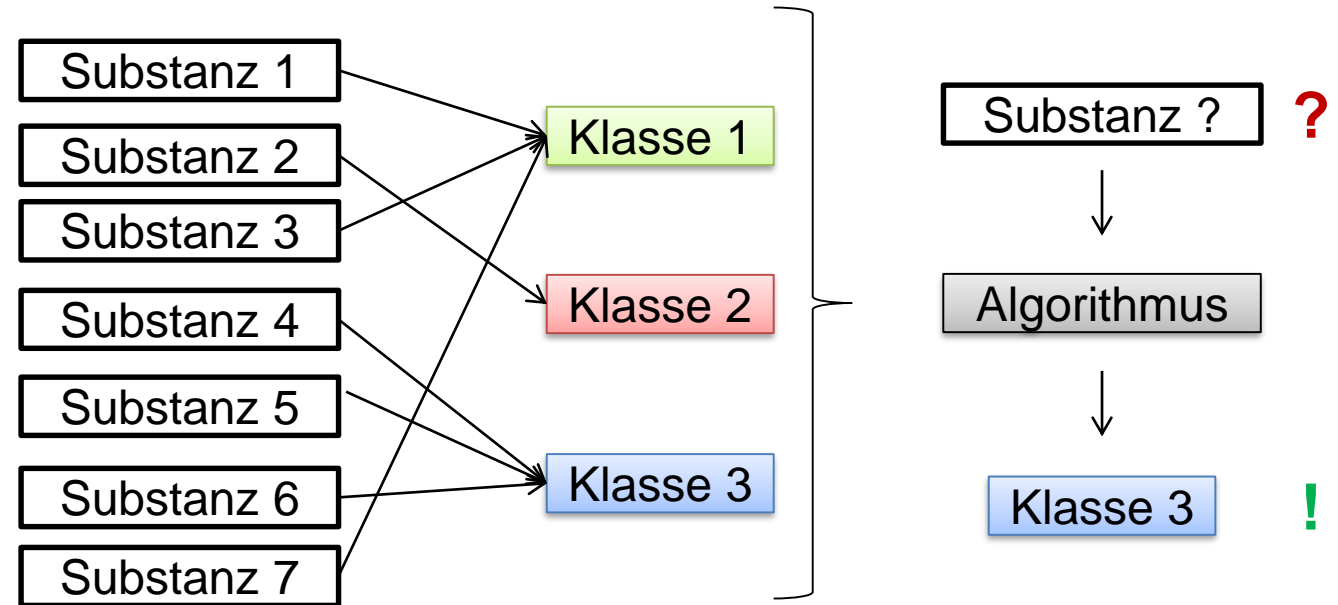
## 2. Training

- Algorithmus wird darauf trainiert, Substanzen nach Vorgaben Klassen zuzuordnen

## 3. Klassifizierung

- Zuverlässig bei Substanzen, auf die der Algorithmus trainiert ist

### Datensatz



# Einteilung der Substanzklassen

Herbizide Pestizide	Lösungsmittel	Mineralöl	Pflanzenöl	Pflanzenstoff	Bakterien
Isoproturon Terbuthylazin Permethrin	Benzaldehyd Diethylether Isopropanol Terpentinersatz	Diesel (Aral) Diesel (HEM) Kerosin Motoröl (2 Takt) Paraffin	Rapsöl Sonnenblumenöl Kokosöl	Chlorophyll A beta-Carotin Lycopin Lutein Pollen <ul style="list-style-type: none"> <li>• Schwarzpappel</li> <li>• Sommer ESP</li> <li>• Sommer GER</li> <li>• Zitterpappel</li> </ul>	B. atrophaeus B. brevis B. subtilis B. thuringiensis E. coli O. urethralis P. polymyxa M. luteus

Chemikalien

Öle

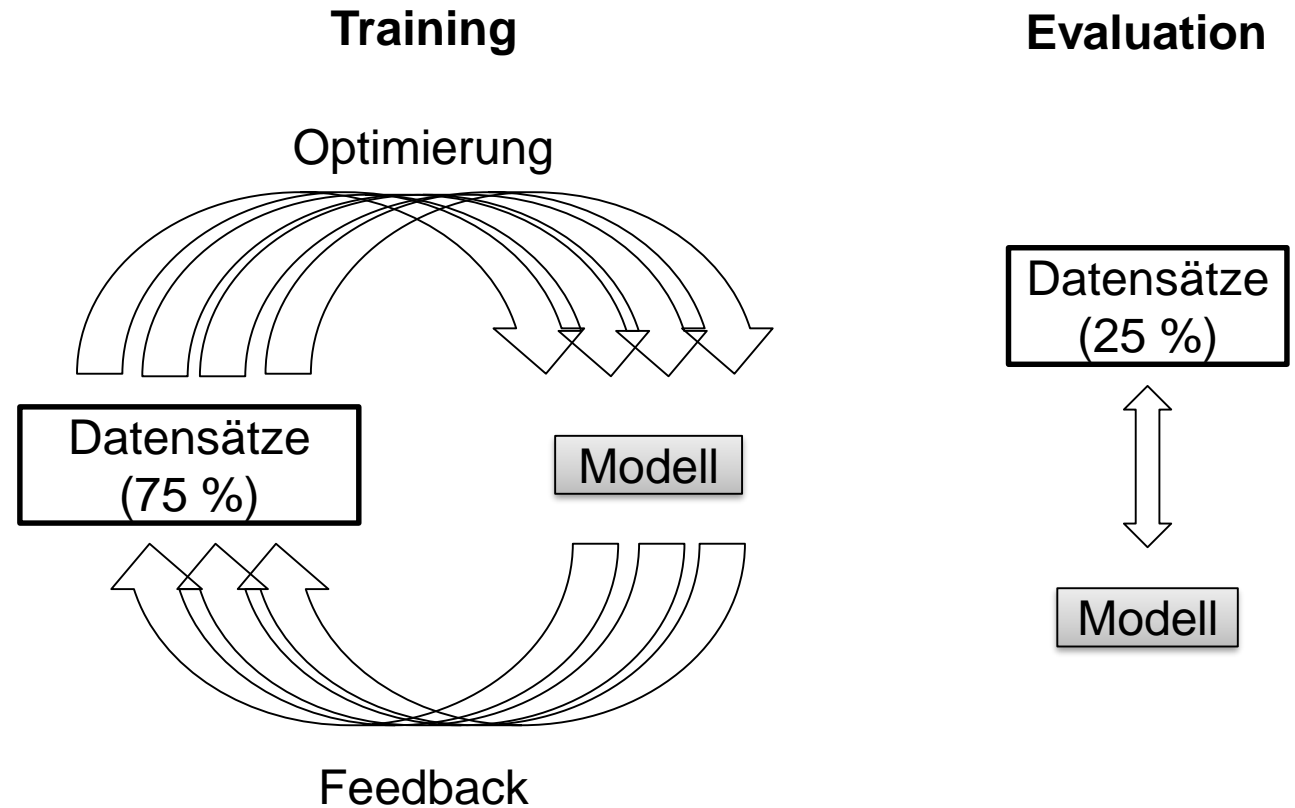
Pflanzen

Bakterien



# Training und Evaluation

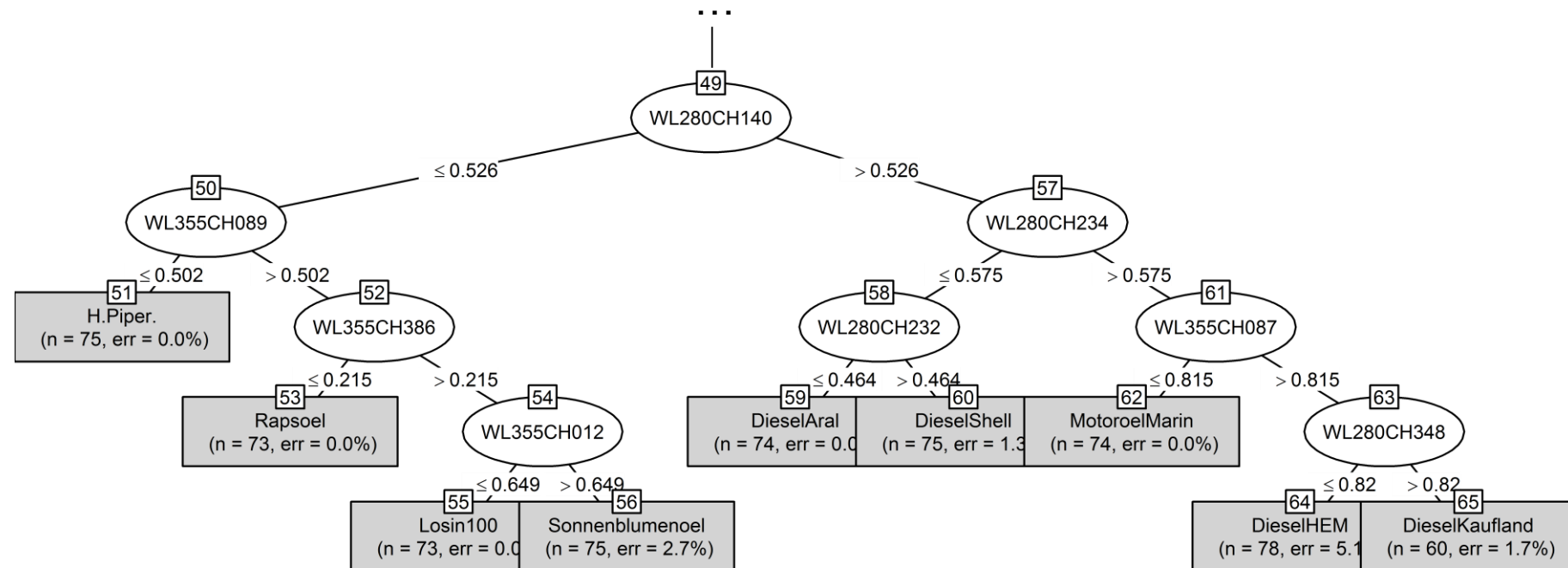
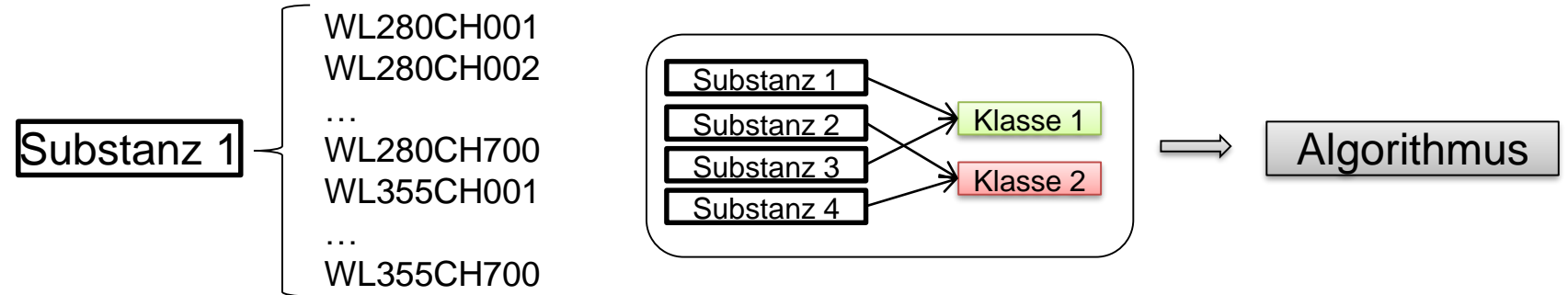
- 8 Substanzen pro Klasse
- 500 Datensätze pro Substanz
- 4000 Datensätze pro Klasse  
(Einige wenige Spektren wurden entfernt)
- 75 % zum Training des Modells  
3000 Datensätze
- 25 % zur Evaluation des Modells  
ca. 1000 Datensätze



# Verwendete Klassifizierungsalgorithmen: Entscheidungsbaum

## Lineare Algorithmen

- Entscheidungsbaum
  - Abfolge von einzelnen Entscheidungen





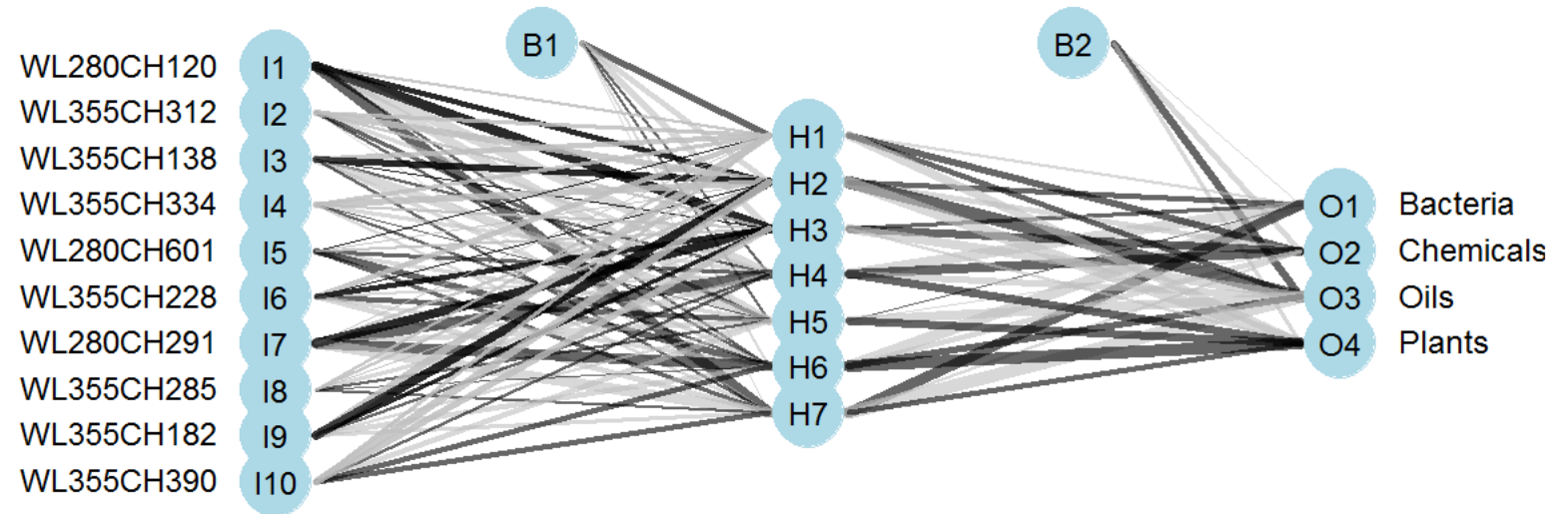
# Verwendete Klassifizierungsalgorithmen: Neuronales Netzwerk

## Lineare Algorithmen

- Entscheidungsbaum

## Nicht lineare Algorithmen

- Neuronales Netzwerk
  - Mustererkennung ohne Abstraktion durch Training
  - Gewichtung einer Auswahl von Kanälen
$$H1 = 0.2 I1 - 0.3 I2 + 0.4 I3 \dots$$
  - Störfaktoren verhindern lokales Minimum



I Input  
H Hidden Layer  
O Output  
B Störfaktoren



# Verwendete Klassifizierungsalgorithmen: Support Vector Machines

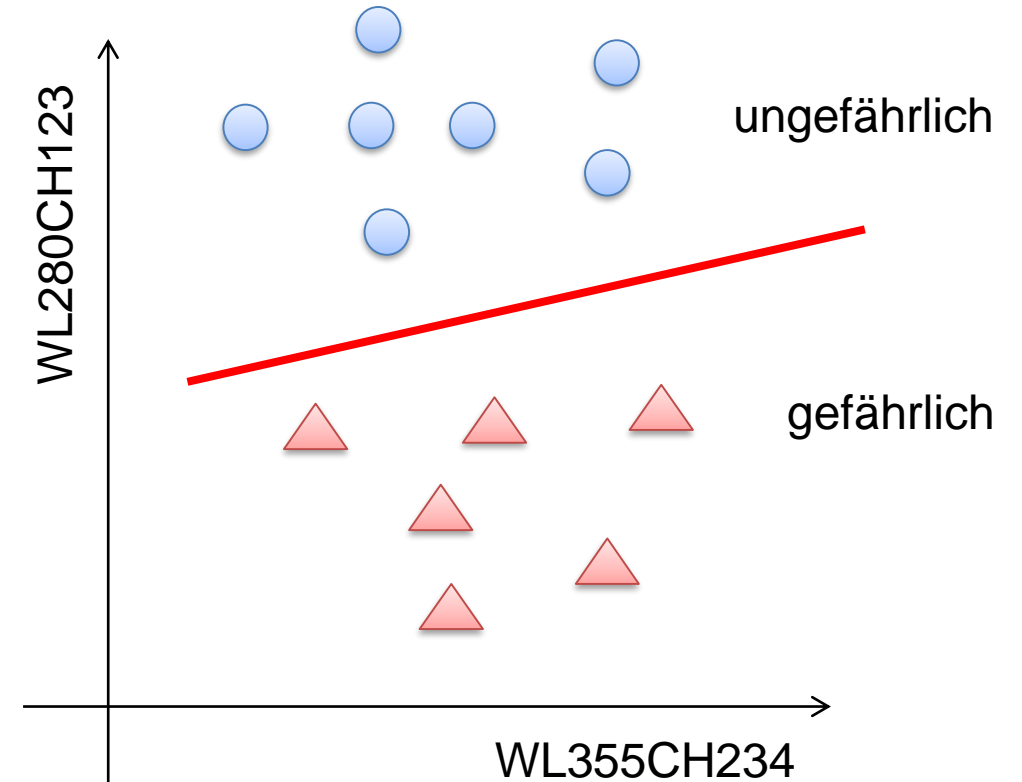
## Lineare Algorithmen

- Entscheidungsbaum

## Nicht lineare Algorithmen

- Neuronales Netzwerk
- Support Vector Machines
  - Jedes Objekt ein Vektor
  - Hyperebene mit maximalem Abstand trennt die Klassen

Bsp:  
2 Kanäle  
2 Klassen  
linear



# Verwendete Klassifizierungsalgorithmen: Support Vector Machines

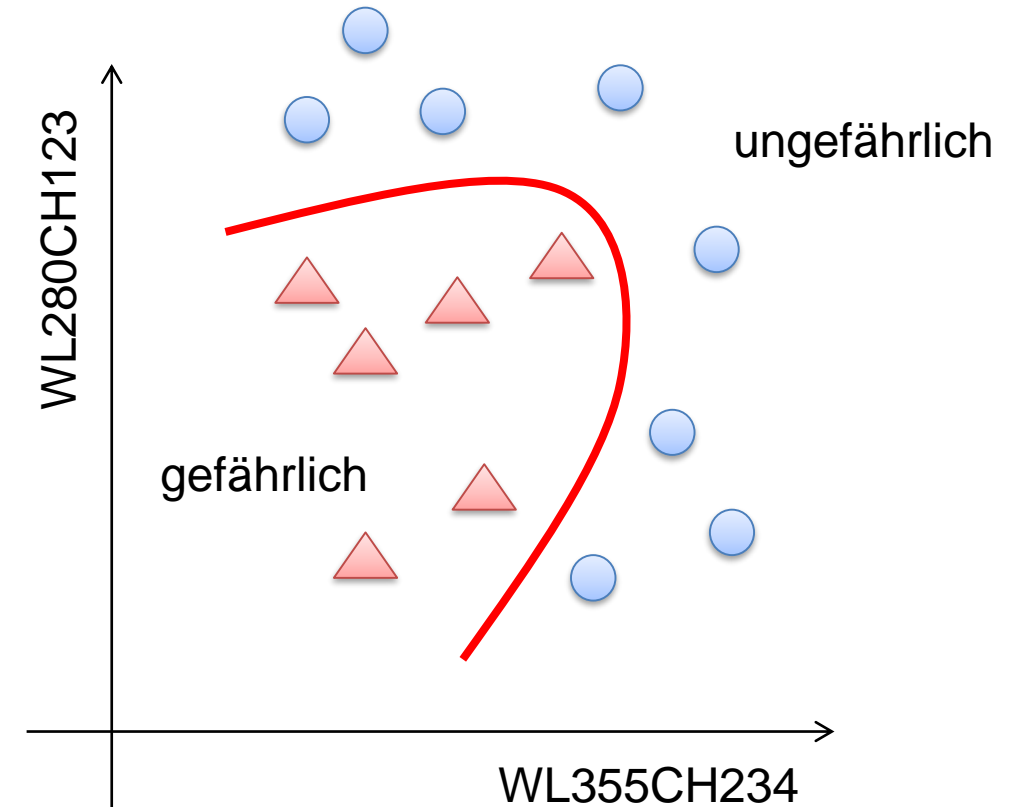
## Lineare Algorithmen

- Entscheidungsbaum

## Nicht lineare Algorithmen

- Neuronales Netzwerk
- Support Vector Machines
  - Jedes Objekt ein Vektor
  - Hyperebene mit maximalem Abstand trennt die Klassen

Bsp:  
2 Kanäle  
2 Klassen  
Nicht linear



# Verwendete Klassifizierungsalgorithmen: Support Vector Machines

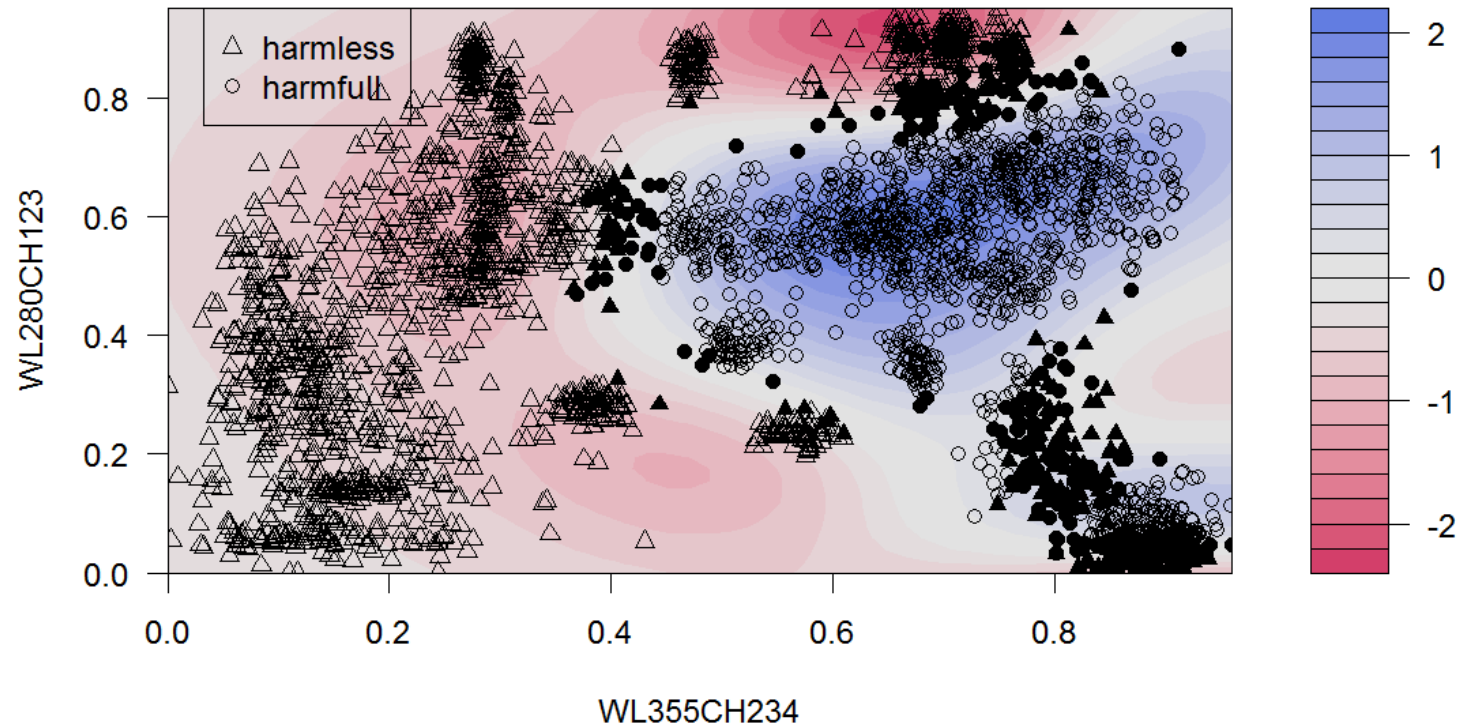
## Lineare Algorithmen

- Entscheidungsbaum

## Nicht lineare Algorithmen

- Neuronales Netzwerk
- Support Vector Machines
  - Jedes Objekt ein Vektor
  - Hyperebene mit maximalem Abstand trennt die Klassen
  - Nur Stützvektoren werden berücksichtigt (*support vectors*)
  - Anzahl der Kanäle bestimmt die Dimension

SVM classification plot



# Vergleich der Klassifizierungsleistungen

- Entscheidungsbaum C5.0  
Accuracy = 0,974

	Vorhersage	Bakterien	Chemikalien	Öle	Pflanzen	Sensitivität	Spezifizität
Referenz	Bakterien	989	2	3	0	0,995	1,0
	Chemikalien	0	966	0	32	0,968	0,977
	Öle	0	1	999	0	0,999	0,999
	Pflanzen	0	65	0	934	0,935	0,989

- Neuronales Netzwerk ANN  
Accuracy = 0,970

	Vorhersage	Bakterien	Chemikalien	Öle	Pflanzen	Sensitivität	Spezifizität
Referenz	Bakterien	991	2	1	0	0,997	1,0
	Chemikalien	0	978	0	20	0,980	0,981
	Öle	0	0	1000	0	1,000	1,000
	Pflanzen	0	55	0	944	0,945	0,993

- Support Vector SVM  
Accuracy = 0,980

	Vorhersage	Bakterien	Chemikalien	Öle	Pflanzen	Sensitivität	Spezifizität
Referenz	Bakterien	989	2	2	1	0,995	0,999
	Chemikalien	1	949	1	47	0,951	0,978
	Öle	2	0	998	0	0,998	0,999
	Pflanzen	0	63	1	935	0,936	0,984





# Ausblick

- Klassifizierungsergebnisse führten zu neuem Experiment mit weniger Kanälen, da hochaufgelöste Fluoreszenzsignale redundante Informationen enthalten  
--> Kurzpuls Lasersystem mit 32 Kanälen
- Demnächst:  
Anwendbarkeit der Klassifizierung auf Raman-Spektren testen
  - Scharfe Peaks
  - Definierte Wellenlängen

